DOI:
10.1016/j.canep.2020.101701

Document Version
Peer reviewed version

Johannesburg Cancer Study (JCS): contribution to knowledge and opportunities arising from 20 years of data collection in an African setting

Authors list

Wenlong Carl Chen,[1,2] *Elvira Singh,[1,3] Mazvita Muchengeti,[1,3] Debbie Bradshaw,[4] Christopher G Mathew,[2,5] Chantal Babb de Villiers,[6] Cathryn M Lewis,[5,7] Tim Waterboer,[8] Robert Newton,[9,10] Freddy Sitas,[4,11,12]

[1] National Cancer Registry, National Health Laboratory Service, Johannesburg, South Africa.

[2] Sydney Brenner Institute for Molecular Bioscience, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa.

[3] School of Public Health, University of the Witwatersrand, Johannesburg, South Africa.

[4] Burden of Disease Research Unit, South African Medical Research Council, Cape Town, South Africa.

[5] Department of Medical and Molecular Genetics, Faculty of Life Sciences and Medicine, King's College London, SE1 9RT, United Kingdom.

[6] Division of Human Genetics, School of Pathology, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa.

[7] Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience, King's College London, SE5 8AF, United Kingdom.

[8] Infections and Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany.

[9] MRC/UVRI and LSHTM Uganda Research Unit, Entebbe, Uganda.

[10] University of York, York, United Kingdom.

[11] Centre for Primary Health Care and Equity, School of Public Health and Community Medicine, University of New South Wales Sydney.

[12] Menzies Centre of Health Policy, School of Public Health, University of Sydney, Australia.

***Corresponding author**

Dr Elvira Singh, elviras@nicd.ac.za

National Cancer Registry, National Health Laboratory Service

1 Modderfontein road,

Sandringham, Johannesburg,

South Africa, 2003

Declaration of Interest

Freddy Sitas, a co-author of this paper, is an Associate Editor of Cancer Epidemiology. The Editor-in-Chief of Cancer Epidemiology managed the editorial process for this manuscript independently and the manuscript was subject to the Journal's usual peer-review process.

Abstract word count: 269

Manuscript word count: 4097

Abstract

The Johannesburg Cancer Study (JCS) aims were to examine whether cancer risk factors identified in Western countries applied to black patients in Johannesburg, South Africa and to understand the impact of HIV on cancer risk, with a view to identifying previously unrecognised HIV associated cancers.

A total of 24 971 black patients with an incident histologically proven (>95%) cancer of any type were enrolled between 1995-2016. Response rates were >90%. Patients provided informed consent, lifestyle and demographic information using a structured questionnaire; 19 351 provided a serum sample and 18 972 a whole blood sample for genomic analyses. This is currently the largest cancer epidemiological biobank in Africa. JCS uses a cancer case-control method; controls being cancer types unrelated to exposures of interest.

Published results show the importance of HIV in several cancers known to be infection associated e.g. Kaposi sarcoma (OR=1683;CI=595-5194) in those with high Kaposi-sarcoma-associated-herpesvirus titres; no effect of HIV on lung or liver cancer-in the latter showing a strong association with HBVDNA, sAg and c positivity (OR=47;CI=21-104). Comparable data to higher-income country studies include lung cancer ORs in relation to smoking (15+g tobacco/day) (OR_{Males}=37;CI=21-67, $OR_{Females}$=18.5;CI=8-45) and associations between alcohol and oesophageal cancer in smokers ($OR_{M\&F}$=4.4;CI=3-6). Relationship between hormonal contraception declined to null 10 or more years after stopping for breast (OR=1.1;CI=0.9-1.4) and cervical cancer (OR1.0;CI=0.8-1.2), and protective effects shown, five or more years after stopping for ovarian (OR=0.6;CI=0.4-1) and endometrial cancer (OR=0.4;CI=0.2-0.9).

Preferential access is based on data requests promoting data pooling, equal collaborative opportunities and enhancement of research capacity in South Africa.

The JCS is a practical and valid design in otherwise logistically difficult settings.

Background

South Africa (SA) has an increasingly high burden of cancer, especially among adults,[1] as leading drivers of carcinogenesis (HIV, smoking, alcohol, reproductive patterns) change in prevalence over time.[2–5] In response to the growing burden of cancer, the Johannesburg Cancer Study (JCS) was established in 1995 at the National Cancer Registry of SA (NCR). Its original aims were to examine whether risk factors identified for cancer in Western countries

applied to black patients in Johannesburg, SA, and to understand the impact of HIV on cancer risk, with a view to identifying previously unrecognised HIV associated cancers. Similar protocols were also used in two other studies. The first, in 1991, was based in Rwanda and was lost with the genocide of 1994.[6] The second was in Uganda, which recruited patients between 1994-1998.[6,7] The JCS aims evolved into measuring the relative importance of known and emerging risk factors for cancer in a local setting and establishing a biobank for ongoing investigations of infectious and genetic drivers of cancer. This is the largest cancer epidemiological study on the continent, with just under 25 000 African cancer patients, providing lifestyle information and blood samples (for serum and DNA).

The JCS was preceded by a medical record review / pilot study between 1992 and 1995 on the association between HIV and cancer among cancer patients attending medical oncology departments of the Charlotte Maxeke Johannesburg Academic Hospital (CMJAH), and its affiliated Radiation Oncology ward of Hillbrow Hospital (ROH).[8] The pilot study demonstrated the feasibility of consecutive patient data collection in local public (oncology) healthcare facilities.

The JCS was launched in 1995 with recruitment commencing using a structured two-page questionnaire at the two sites mentioned above and the haematology and medical wards of Chris Hani Baragwanath Academic Hospital (CHBAH). CHBAH refers most cancer patients eligible for radiation therapy and medical oncology to ROH or CMJAH so recruitment at CHBAH was terminated in 2001. ROH relocated to CMJAH in 2005 and patient recruitment at the CMJAH was terminated in April 2016. Patients undergoing cancer treatment generally require pathological confirmation of their cancer diagnosis, hence the proportion of cancer cases verified histologically was >95% (Table 1), substantially higher than in similar studies in Uganda and Rwanda.[7]

Recruitment took place in a consecutive manner at the medical oncology or radiation therapy ward waiting rooms. Eligible patients were identified through clinic daily logbooks, approached by trained research oncology nurses and invited to participate in the JCS. Participants provided signed informed consent or witnessed oral consent. Cancer counselling was available to participants by our oncology research nurses. All interviews were done in private rooms in the preferred language of the patient, mainly Zulu, Sotho, related dialects or English.

Inclusion /Exclusion Criteria /Case definition

Incident cancer patients self-identified as black African, aged 18 and older, able to provide informed consent to participate, able to speak one of the main languages in SA and optionally able to provide two blood samples were included in the study. Persons with treatment naïve, incident cancer were identified by reviewing medical records and preliminary patient discussions. Diagnosis of primary site of cancer and morphology were determined by review of medical records and pathology results of the tumour biopsies. Self-identified non-resident South African patients were excluded.

Data collected

Demographic data collection

The JCS collected diagnostic, lifestyle and demographic information using a two-page questionnaire. These included tumour topography of primary site and morphology (from medical records), lifestyle and demographic questions, date of birth and interview, place of birth and current residence, highest educational attainment, type of fuel used for heating and cooking, smoking (current, former and never, dates started and stopped, types and amounts smoked), snuff use (never, past, current), alcohol status (initially type of alcohol and frequency consumed, later changed to type, frequency and amount consumed), history of self-reported high blood pressure and diabetes, occupation (type of work and industry), number of sexual partners, and main language of mother and father. For women, separate questions included age at first and last child, number of live children and miscarriages, and frequency and type of hormonal contraception used (oral or injectable). In 2004 questions on knowledge about antiretrovirals (ARV) and HIV status were added.

Questionnaires and pathology reports were returned to the NCR for dual data capture and data quality control. Cancer topography and morphology was coded by NCR trained nosologists using ICDO-3 and into diagnostic categories using IARC ICDO-3 to ICD-10 conversion rules.

Blood sample collection

Venous blood samples were collected from participants starting in 1996 in a BD vacutainer® serum separation tube. This was processed, usually within 24 hours by centrifuging at 1000-2000g for 10 min in a refrigerated centrifuge and dividing the serum into ~4 aliquots in NUNC cryotubes and initially stored at -25°C. Serum was moved into -80°C freezers in 2011. The serum collected were of adequate quality and quantity for serological screenings of specific antibodies against multiple pathogens of carcinogenic risk to humans outlined by the IARC Monographs Volume 100B (such as HIV, KSHV, HPV).[9-12]

Participants were asked by the study interviewers if they wished to have an HIV test performed, with results returned to them. If affirmative, interviewers drew a separate (non-study) blood sample for testing using the pathology laboratory's routine system. Interviewers were also trained in pre- and post-HIV counselling and report-back was discussed with the patient during their next appointment and results incorporated as part of their routine care.

Blood collection for DNA isolation commenced in 1997 in one BD® 4ml EDTA vacutainer. Vacutainers were stored in -30°C for future DNA isolation. We demonstrated recently that whole blood samples stored (between 2-19 years) at -30°C yielded sufficient DNA with adequate quality for downstream genetic analysis such as genotyping genetic variants on multiple platforms and DNA sequencing.[13,14]

Study recruitment continued at ~1100 cases per annum. At the end of recruitment in 2016, 24 971 patients were enrolled, 19 351 had a serum sample and 18 972 had a whole blood sample collected. Between 1998-2001 we also recruited 1173 patients presenting with cardiovascular diseases as a cancer-negative control group.

Ethics

The JCS has University of the Witwatersrand (Wits) Human Research Ethics Committee (Medical) (HREC) clearance (M981119, M040445, M090361, M140271, M1606103, M120117). Participants provided witnessed consent to a once-off interview and optional blood draw and to have their information and samples stored anonymously for any future infection or inherited risk factor investigations for cancer, with the understanding that no research would be done without the approval of the Wits HREC as well as the applicable HREC at host institutions.

Data Resource Use:

Study design:

The study design we used to perform several analyses is akin to a hospital-based case-control study. For each study (hypothesis) controls were selected on a case-by-case basis by first defining the appropriate cases and then choosing sex and age range matched participants with cancers unrelated to the main exposures of interest under investigation (see illustration in Table 3). So, for example, in a study on the effects of smoking on lung cancer we excluded all those participants with cancers that were suspected or known to be associated with smoking.[15] We identified these by reviewing IARC Monographs on Carcinogenicity in the first instance, and if these were out of date we referred to informative reviews such Schottenfeld and Fraumeni's textbook on cancer epidemiology (3rd and 4th editions), and by literature searches to identify other influential emerging studies. This has proven to be practical solution under an otherwise logistically difficult setting. The analyses used for selected lifestyle and serological data comprised calculations of case-control odds ratios adjusting for various known confounders.

A choice of similarly-ill cancer controls, compared to the cases under investigation has the advantage of minimising referral bias (assuming all cancers are referred through similar pathways and catchment areas), interviewer bias (interviewers were never sure which hypothesis is being tested at any time) and recall bias (participants are both similarly sick, reducing potential recall biases of prior exposures between sick vs. healthy individuals). We minimised obvious biases by performing sensitivity analyses in which we removed each main cancer type from the pool of controls and recalculated odds ratios using the remaining controls,[16] or by calculating the heterogeneity of exposure prevalences by cancer type among controls.[9] We interviewed 1173 patients with cardiovascular disease which served as an additional negative control group (in the case of an HIV and cancer study where no association between immune suppression and cardiovascular disease was suspected),[10] and as an additional (positive) case group, when investigating smoking related cancers.[15]

Evolution of risk factors:

Table 1 illustrates the changes in prevalence of some of the key exposures in two periods over the 21-year life-course of this study, 1995-2004 and 2005-2016. For certain exposures such as tobacco smoking, alcohol consumption, hormonal risk factors and HIV seropositivity, the

prevalence was estimated by excluding cancers with known risk indications for the specific exposures to reduce sampling biases.

One of the most important changes in exposures has been HIV seropositivity in cancers unrelated to known infectious agents, from 6.6% (in men) and 8.8% (in women) in the first period to 15.4% and 21.6%, respectively in the second period. This increase in the background HIV prevalence had a profound effect on the distribution of cancer in this population, with Kaposi sarcoma ranking first in men and third in women (previously ranked fifth in men and eighth in women).

Smoking prevalence in men and women remained about the same but the median number of cigarettes dropped slightly. Alcohol consumption has decreased in women over the two periods. The use of coal and anthracite as fuel sources decreased, while the prevalence of electricity use increased, in keeping with government initiatives to electrify Soweto (Johannesburg's former satellite township) and other formerly black and under-invested areas.

Table 1 also shows the evolution of the top three cancer types changing in men, in order, from oesophagus, prostate and lung in 1995-2006 to Kaposi Sarcoma, oropharyngeal and lung in 2005-2016. Changes in cancer relative frequency were also observed in women, with cervix, breast and oesophageal cancer being the top three cancers in 1995-2004 to breast, cervix and Kaposi Sarcoma in 2005-2016. Similar increases in rankings have been seen in the rarer infection related cancers such as the lymphomas, eye cancers and some genital cancers; these changes are also observed in the NCR national surveillance data.[17]

[Insert table 1]

Table 2 illustrates examples of the number of participants by cancer type and choice of potential controls in four scenarios, investigating smoking, alcohol, infection and hormonal contraception related cancers.

[Insert table 2]

<u>Documentation of cancer risks in an urban African population:</u>

Table 3 summarises some of the main results from the JCS. The first paper from the study documented Kaposi Sarcoma odds ratios of 1 683 in those who were HIV infected and had high titres of Kaposi's sarcoma-associated herpesvirus (KSHV), compared to about 12-fold risks in those who were not infected with HIV. This illustrates that against a background of KSHV the causal association between HIV and Kaposi Sarcoma, one of the first syndromes to be defined by the CDC as an AIDS related cancer.[18] This important locally derived information was cited by the "Durban Declaration" [19] in response to South African government denial of the role HIV played in causing AIDS.

The risk for hepatocellular carcinoma (HCC) increases significantly in the presence of hepatitis B viral infection and HIV co-infection, however the HIV alone does not appear to be a risk factor for HCC in the South African black population.[11] This may be suggestive of impaired innate and adaptive immune response in HIV co-infection leads to an reduction of

inflammation related to immune-mediated clearance of HBV-infected hepatocytes.[20,21] Thus, effective control of hepatitis B infection through the use of vaccines at infancy may be an effective method for reducing risk of HCC in a population burdened with HIV.

The JCS has documented the local effects of some of the main drivers of cancer risk such as HIV infection, causing increases in AIDS-related (viz. Kaposi Sarcoma, cervix, lymphoma) and other infection related cancers, including an increased risk for squamous cell skin cancer but no increased risk in relation to hepatocellular and lung cancer.[10] One peculiarity of these data was that the risks for e.g. Kaposi Sarcoma and HIV, were one or two orders of magnitude lower than what was found in Western countries (~20-50 vs. >1000). This was explained by the (endemic) background incidence of KS in (South) Africa being much higher than in developed countries.[10] Other contributions include data to a 10-country collaboration showing a null effect of HPV on oesophageal cancer. [22]

The JCS also documented expected risks of smoking and lung cancer (and several other cancers) among heavy smoking black males (15g + tobacco per day) of ORs of about 24-37-fold, which is similar to what was found in e.g. the British Doctors study of RR>16.9.[23] Previous studies from Africa had very few people who were heavy smokers hence the higher risks in the JCS compared to other African studies and lower risks when compared to UK studies.[24] In conjunction with smoking we documented the increased risks in relation to alcohol consumption in upper aerodigestive cancers, particularly among smokers.[25]

In SA, the proportion of women using injectable contraceptives was three times higher than those using oral contraceptive pills.[26] We showed that recent use of especially injectable hormonal contraception is associated with a modest increase in the risks of breast and cervical cancer, and reduced risks for ovarian cancer. Relationship between hormonal contraception declined to null 10 or more years after stopping for breast (OR=1.1;CI=0.9-1.4) and cervical cancer (OR1.0;CI=0.8-1.2), and protective effects shown, five or more years after stopping for ovarian (OR=0.6;CI=0.4-1) and endometrial cancer (OR=0.4;CI=0.2-0.9), in keeping with international studies.[16,27]

[Insert table 3]

<u>Contribution to international and local collaborations:</u>

The JCS has enjoyed success with its international and local collaborations and training of students. Several South African Medicine interns, MSc and PhD students have used the data for a range of projects. The JCS has contributed data to the International Collaboration of Epidemiological studies on breast, endometrial and ovarian cancers,[27,28] International Collaboration on cervical cancer[29] and data and serum samples were pooled in an international study on HPV and oesophageal cancer (InterSCOPE Study).[22] In recent years, JCS contributed DNA samples to the Men of African Descent and Carcinoma of the Prostate (MADCaP) consortium investigating the genetic and epidemiological risk for prostate cancer in resident African populations.[30] DNA from cases of oesophageal squamous cell carcinoma was used to demonstrate association of genetic variants in the *CHEK2* gene with this cancer.[14] Local collaborations included a study on hepatocellular carcinoma[11], knowledge of HIV status at cancer diagnosis[31] and versions of the questionnaire were used by colleagues in two other

local settings.[32,33] Curent collaborations include an extensive assessment of genomic effects on prostate[34], cervical, breast and oesophageal cancer, and the role of key lifestyles and 18 infectious agents ('onco-agents', 10 known and eight suspected to increase cancer risk) in association with several cancer types.[35]

Strengths:

The successes of the JCS in a resource constrained environment can be attributed to a number of factors. The design of the questionnaire was simple, short and practical allowing for sick patients to be interviewed while awaiting consultation in radiation and medical oncology clinics. Highly trained research nurses administered the interviews which required minimal input from busy clinic staff. Patients found added value in participation as the interviewers provided counselling on their cancer diagnosis and the path that lay ahead for them. The study was conducted over many years, so the interviewers were integrated into the clinical team. This enhanced the clinical service offered through nurse interviewer-patient interaction, occasionally incorporating HIV pre- and post-test counselling.

In places where there are tertiary hospitals with reasonably good cancer treatment and diagnostic facilities, and where cohort studies are impractical, the JCS offers a pragmatic methodology to obtain information on cancer risk factors in otherwise difficult study settings. In this way it was possible to obtain up-to-date information on the relative importance of key risk factors in a setting and develop local capacity in cancer epidemiology and genomics. Many of the current, relevant and common cancer risk factors were fitted into a 2-page questionnaire. It is possible to expand this questionnaire into three pages using[31] a common set of questions and add a fourth page focusing on local risk factors,[5] and (where practical) extend the recruitment to cancer-free spouses of cancer patients to obtain an additional non-cancer control group.[36] We did attempt interviewing spouses but this proved logistically difficult as many spouses lived too far away to visit their sick partners.

JCS has recruited cancer patients with high histological verification rates in a continuous fashion over 20 years. A carefully annotated collection of biobanked samples with lifestyle data is now available. Response rates for the study have been excellent, over 90%.[10] Questionnaire completion rates were also high. Using cancer controls was a practical design which also minimizes most of the known case control study biases. Consistent questions were asked for most exposures over the course of the study allowing for historical comparisons. Reassuringly, the prevalence of key exposures in the controls studied resembles the background prevalence in the population.

High quality DNA obtained from peripheral blood samples are usable for DNA sequencing studies to elucidate the prevalence of rare hereditary genetic risk factors (e.g. *BRCA1* or *BRCA2* for breast cancer) or genotyping of small numbers of candidate genes to genome-wide association studies (with appropriately matched external population-control data) to identify common genetic risk factors.[13,14]

<u>Limitations:</u>

Given the retrospective nature of the study retrospective recall of past exposures limits which questions were asked. For this reason, dietary questions were avoided. Self-reporting of various pills and treatments can be problematic in low health literacy populations and may also be subject to some recall biases. Pre-cancer symptoms may have an effect on lifestyle behaviour, for example early symptoms of lung cancer e.g. coughing may cause some to stop smoking as a result of the illness. While we have controlled some of this by asking start and stop dates of smoking, this was not controlled for in several other questions (e.g. for snuff use) so some residual confounding by indication may still be possible. Of course, as new cancer – exposure associations are discovered post analysis, some control participants would then be deemed as wrongly classified. For example, colorectal cancer was only recently deduced to be associated with smoking, breast cancer with alcohol consumption[37] and recent meta-analyses suggest a protective effect of alcohol on lymphoma (albeit in all cases the relative risks (RR) are relatively small).[38] Such new evidence will inform future case control selection, but because each of these cancers is only a small fraction of all the cancer types comprising the control group, their overall influence on the odds ratios is minimised.

Aside from providing interviewers to each clinic, this study was managed on a minimal budget, offered no administrative/infrastructure hospitals costs. As such the study depended on in-kind support from the hospital, so this study was in constant competition with other better-funded studies or trials. JCS also depended on good relationships and goodwill from the heads of hospital departments and associated staff. Although relationship building was successful, staff turnover meant that this was an ongoing process. Staff turnover within the study meant that skilled personnel were lost to the study and training was required once again.

Although patient recruitment finished in 2016, there are still ongoing associated costs. These include the cost of long-term storage of samples, DNA extraction, ethics report-backs, data curatorship and publication expenses. However, the data and sample repository from this study remains an invaluable asset to cancer researchers, particularly given the current interest in the genetic risk factors for cancer in African ancestry patients, and efforts are being made to maintain this asset for future studies.

A further challenge is that cohort studies (rather than case control) are now increasingly considered the gold standard in epidemiology. However, they are very expensive, especially if biobanking is involved, and difficult to conduct in resource constrained environments due to loss to follow up.

Assuming an adult cancer incidence of 188 per 100 000, a cohort of ~584 000 adults would be needed to yield ~1100 cancer patients (recruited by the JCS annually).[39] Therefore, a well-designed, strictly managed case-control study was chosen to be the study design of choice in this setting.[40-43]

Data Resource Access:

Access to data from the JCS is managed through a data access agreement with the National Cancer Registry of South Africa. In brief: detailed protocols are submitted and reviewed by

internal/expert research committees. Data are released to the investigators once protocols are approved and ethical clearance obtained, and data transfer agreements and memoranda of understanding are signed. The JCS database contains only de-identified data linked by a study number to blood and serum samples. Ethical approval for using DNA and associated data from this repository is required for each new study. Samples and data shared with researchers external to the NCR are required to have a Material Transfer Agreement (MTA) in place. We have previously obtained ethical permission to send DNA and serum for microarray DNA analyses and multiplex serum analyses to the United Kingdom and Germany with standard MTAs from the Wits HREC. Cost-recovery principles apply. Preference is placed on proposals which promote data pooling, equal collaborative opportunities, and enhancement of research capacity in SA. Sharing of anonymised genetic data with the scientific community will be done via resources such as the European Genome-phenome Archive after approval by the Wits HREC and an approved data access committee.

Funding:

Original seed funding was provided by Valerie Beral (Cancer Research UK, Cancer Epidemiology Unit). Subsequently the study infrastructure was funded by the South African Medical Research Council, the National Health Laboratory Service and the Cancer Association of South Africa. Currently, the study has been supported by the South African Medical Research Council (with funds received from the South African National Department of Health) and the UK Medical Research Council (with funds from the UK Government's Newton Fund) (MRC-RFA-SHIP 01-2015). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of the South African Medical Research Council or the South African National Department of Health or the UK Medical Research Council from the UK Government's Newton Fund.

Acknowledgements:

This study was conceived in 1990 by Professor Valerie Beral (Cancer Research UK) in discussion with FS. The questionnaire was devised by Valerie Beral, FS and RN. Several clinicians facilitated the study and recruitment. These include Alan Fleming, Bernard Donde, David Saffer, Martin Hale, Moosa Patel, Myles Connor, Paul Ruff, Pradeep Rowji, Ranjan Sur, and Victor Levin. Several NCR staff worked on the administrative and scientific aspects of the study: Edith Ratshikopha, Gavin Fleming, Hellen Mathabatha, Henri Carrara, Lara Stein, Lettie Bester, Margaret Urban, Funani Mgnomezulu, Janet Knuppel, Jitesh Madhoo, Nokuzola Mqoqi and Rosana Pacella-Norman. Interviewers included: Gloria Mokwatle, Meriam Ramela, Patricia Rapoho, Pheladi Kale, Thandi Zwane and Tsakani Mhlongo.

References:

1. Bradshaw D, Dorrington RE, Sitas F. The level of mortality in South Africa in 1985--

what does it tell us about health? *S Afr Med J*. 1992;**82**(4):237–40.

2. Parkin DM, Sitas F, Chirenje M, Stein L, Abratt R, Wabinga H. Part I: Cancer in Indigenous Africans--burden, distribution, and trends. *Lancet Oncol*. 2008;**9**(7):683–92.

3. Sitas F, Parkin DM, Chirenje M, Stein L, Abratt R, Wabinga H. Part II: Cancer in Indigenous Africans--causes and control. *Lancet Oncol*. 2008;**9**(8):786–95.

4. Mayosi BM, Flisher AJ, Lalloo UG, Sitas F, Tollman SM, Bradshaw D. The burden of non-communicable diseases in South Africa. *Lancet (London, England)*. 2009;**374**(9693):934–47.

5. Mayosi BM, Lawn JE, Niekerk A van, et al. Health in South Africa: changes and challenges since 2009. *Lancet (London, England)*. 2012;**380**(9858):2029–43.

6. Newton R, Ngilimana P-J, Grulich A, et al. Cancer in Rwanda. *Int J Cancer*. 1996;**66**(1):75–81.

7. Beral V, Newton R. Overview of the epidemiology of immunodeficiency-associated cancers. *J Natl Cancer Inst Monogr*. Narnia; 1998;**1998**(23):1–6.

8. Sitas F, Bezwoda WR, Levin V, et al. Association between human immunodeficiency virus type 1 infection and cancer in the black population of Johannesburg and Soweto, South Africa. *Br J Cancer*. 1997;**75**(11):1704–7.

9. Sitas F, Carrara H, Beral V, et al. Antibodies against human herpesvirus 8 in black South African patients with cancer. *N Engl J Med*. 1999;**340**(24):1863–71.

10. Stein L, Urban MI, O'Connell D, et al. The spectrum of human immunodeficiency virus-associated cancers in a South African black population: results from a case-control study, 1995-2004. *Int J Cancer*. 2008;**122**(10):2260–5.

11. Mak D, Babb de Villiers C, Chasela C, Urban MI, Kramvis A. Analysis of risk factors associated with hepatocellular carcinoma in black South Africans: 2000-2012. *PLoS One*. 2018;**13**(5):e0196057.

12. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Biological Agents. Vol 100B, IARC, Lyon, 2009.

13. Chen WC, Kerr R, May A, et al. The Integrity and Yield of Genomic DNA Isolated from Whole Blood Following Long-Term Storage at −30°C. *Biopreserv Biobank*. 2018;**16**(2):106–113.

14. Chen WC, Bye H, Matejcic M, et al. Association of genetic variants in CHEK2 with oesophageal squamous cell carcinoma in the South African Black population. *Carcinogenesis*. 2019;**40**(4):513-520.

15. Stein L, Urban MI, Weber M, et al. Effects of tobacco smoking on cancer and cardiovascular disease in urban black South Africans. *Br J Cancer*. 2008;**98**(9):1586–92.

16. Urban M, Banks E, Egger S, et al. Injectable and oral contraceptive use and cancers of the breast, cervix, ovary, and endometrium in black South African women: case-

control study. *PLoS Med*. 2012;**9**(3):e1001182.

17. National Cancer Registry of South Africa. Cancer in South Africa 2014 Full report. (2014). http://www.nicd.ac.za/wp-content/uploads/2019/12/2014-NCR-tables.pdf, accessed [25 February 2020].

18. Hengge UR, Ruzicka T, Tyring SK, et al. Update on Kaposi's Sarcoma and Other HHV8 Associated Diseases. Part 1: Epidemiology, Environmental Predispositions, Clinical Manifestations, and Therapy. *Lancet Infect Dis.* 2002;**2**(5):281-292.

19. The Durban Declaration. *Nature*. 2000;**406**(6791):15–6.

20. Guidotti LG, Chisari FV. Immunobiology and pathogenesis of viral hepatitis. Annu Rev Pathol Mech Dis. 2006;**1**:23–61.

21. Puoti M, Torti C, Bruno R, Filice G, Carosi G. Natural history of chronic hepatitis B in co-infected patients. Journal of hepatology. 2006;**44**:S65–S70.

22. Sitas F, Egger S, Urban MI, et al. InterSCOPE study: Associations between esophageal squamous cell carcinoma and human papillomavirus serological markers. *J Natl Cancer Inst*. 2012;**104**(2):147–58.

23. Doll R, Peto R, Boreham J, Sutherland I. Mortality from cancer in relation to smoking: 50 years observations on British doctors. Br J Cancer. 2005;**92**(3):426-9.

24. Parkin DM, International Agency for Research on Cancer. Cancer in Africa : epidemiology and prevention. IARC Press; 2003.

25. Pacella-Norman R, Urban MI, Sitas F, et al. Risk factors for oesophageal, lung, oral and laryngeal cancers in black South Africans. *Br J Cancer*. 2002;**86**(11):1751–6.

26. United Nations, Department of Economic and Social Affairs, Population Division (2015). Trends in Contraceptive Use Worldwide 2015 (ST/ESA/SER.A/349).

27. Collaborative Group on Epidemiological Studies on Endometrial Cancer. Endometrial cancer and oral contraceptives: an individual participant meta-analysis of 27 276 women with endometrial cancer from 36 epidemiological studies. *Lancet Oncol*. 2015;**16**(9):1061–1070.

28. Collaborative Group on Epidemiological Studies of Ovarian Cancer, Beral V, Gaitskell K, et al. Ovarian cancer and smoking: individual participant meta-analysis including 28,114 women with ovarian cancer from 51 epidemiological studies. *Lancet Oncol*. 2012;**13**(9):946–56.

29. International Collaboration of Epidemiological Studies of Cervical Cancer. Cervical carcinoma and reproductive factors: collaborative reanalysis of individual data on 16,563 women with cervical carcinoma and 33,542 women without cervical carcinoma from 25 epidemiological studies. *Int J cancer*. 2006;**119**(5):1108–24.

30. Andrews C, Fortier B, Hayward A, et al. Development, Evaluation, and Implementation of a Pan-African Cancer Research Network: Men of African Descent and Carcinoma of the Prostate. *J Glob Oncol*. 2018;(4):1–14.

31. Sengayi M, Babb C, Egger M, Urban MI. HIV testing and burden of HIV infection in

black cancer patients in Johannesburg, South Africa: a cross-sectional study. *BMC Cancer*. 2015;**15**:144.

32. Sewram V, Sitas F, O'Connell D, Myers J. Tobacco and alcohol as risk factors for oesophageal cancer in a high incidence area in South Africa. *Cancer Epidemiol*. 2016;**41**:113–121.

33. Mzileni O, Sitas F, Steyn K, Carrara H, Bekker P. Lung cancer, tobacco, and environmental factors in the African population of the Northern Province, South Africa. *Tob Control.*1999 Winter;**8**(4):398-491.

34. Odiaka E, Lounsbury DW, Jalloh M, et al. Effective Project Management of a Pan-African Cancer Research Network: Men of African Descent and Carcinoma of the Prostate (MADCaP). J Glob Oncol. 2018;**4**:1-12.

35. SAMRC. Evolving Risk Factors for Cancers in African Population (ERICA-SA). http://www.mrc.ac.za/intramural-research-units/evolving-risk-factors-cancers-african-populations-erica-sa , accessed: 5th February 2020.

36. Sitas F, Yap S, Egger S, et al. The Cancer, Lifestyle and Evaluation of Risk Study (CLEAR): Rationale and design of an unmatched "case-spouse control" study of over 10,000 participants in New South Wales, Australia. *Cancer Epidemiol*. 2015;**39**(3):414–423.

37. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Part E: Personal habits and indoor combustions. Vol 100E, IARC, Lyon, 2009.

38. Psaltopoulou T, Sergentanis TN, Ntanasis-Stathopoulos I, Tzanninis IG, Tsilimigras DI, Dimopoulos MA. Alcohol consumption and risk of hematological malignancies: A meta-analysis of prospective studies. *Int J Cancer*. 2018;**143**(3):486-495.

39. Ferlay J, Ervik M, Lam F, Colombet M, Mery L, Piñeros M, Znaor A, Soerjomataram I, Bray F (2018). Global Cancer Observatory: Cancer Today. Lyon, France: International Agency for Research on Cancer. Available from: https://gco.iarc.fr/today, accessed [25 February 2020].

40. Linet MS, Brookmeye R. Use of Cancer Controls in Case-Control Cancer Studies. *Am J Epidemiol.* 1987;**125**(1):1-11.

41. Miettinen OS. The "case-control" study: Valid selection of subjects. *J Chron Dis.* 1985;**38**(7):543-548.

42. Pearce N, Checkoway H. Case-control Studies Using Other Diseases as Controls: Problems of Excluding Exposure-Related Diseases. *Am J Epidemiol.* 1988;**127**(4):851-856.

43. Verhage BAJ, Aben KKH, Straatman H, et al. Spouse controls in family case-control studies: a methodological consideration. *Familial Cancer.* 2003;**2**:101-108.

Table 1. Basic demographic and lifestyle characteristics of cancer patients in the Johannesburg cancer study.

	1995 – 2004 (n=9 530)[#]	2005-2016 (n= 15 441)[#]
Top three cancer types (% of total) – Males (M)	Oesophagus (17.0%), Prostate (11.7%), Lung (10.5%)	Kaposi Sarcoma (13.1%), Oral cavity and pharynx (11.3%), Lung (10.2%)
Top three cancer types (% of total) – Females (F)	Cervix (33.4%), Breast (24.8%), Oesophagus (5.8%)	Breast (33.4%), Cervix (30.7%), Kaposi Sarcoma (4.8%)
Cancers histologically verified	97.4%	98.9%
Cancers ill defined	3.0%	3.5%
% Female	63.7%	68.4%
Median age – years (IQR)	52 (42-62)	51 (41-60)
Age-group: 18-24 years	2.8%	2.0%
25-54 years	53.2%	59.1%
55-74 years	38.4%	37.9%
75+ years	5.6%	1.0%
Urban place of birth	46.6%	49.0%
Urban place of residence	81.7%	91.5%
Secondary school leavers	8.1%	20.5%
Using electricity to cook (now)	69.5%	84.2%
Using coal + anthracite to cook (now)	11.1%	2.7%
Using electricity for heating (now)	65.0%	65.7%
Using coal + anthracite for heating (now)	16.8%	11.7%
Smoker (M) – current smokers[α]	31.4%	30.1%
Smoker (F) – current smokers[α]	5.1%	4.9%
Median number of cigs/day (M)[α]	9	7
Median number of cigs / day (F)[α]	5	4
≥ Moderate alcohol drinkers (M) (>200g per week)[β]	39.2%	39.9%
≥ Moderate alcohol drinkers (F) (>200g per week)[β]	12.5%	9.4%
HIV positive (M)[χ]	6.6%	15.4%
HIV positive (F)[χ]	8.8%	21.6%
Currently using oral contraception only (F only 18-44 years)[δ]	5.3%	3.3%
Currently using injectable contraception only (F only 18-44 year)[δ]	16.4%	19.8%
Median age at first childbirth (15-54 years as childbearing age) (IQR)[δ]	21 (19-24)	20 (18-22)
Median number of children (F)	3	3
Median self-reported sexual partners	4	4
Language Zulu	24.3%	25.3%
Language Sesotho	19.5%	17.9%
Language Tswana	19.1%	17.5%
Language Xhosa	13.3%	12.1%

Note these figures are illustrative, the percentages are calculated per each study period - comprising crude estimates in both potential cases and controls - see Table 2.

[#]Total number of cancer patients interviewed and recruited within each study period.

[α]Smoking behaviour was estimated with the exclusion of smoking-related cancers.

[β]Alcohol consumption was estimated with the exclusion of alcohol-related cancers.

[χ]HIV prevalence was estimated in cancers unrelated to known infectious agents.

[δ]Contraception use were estimated in female cancers unrelated to reproductive or hormonal factors.

Table 2. Potential controls selection for cancer types in four scenarios: investigating smoking, alcohol, infection and hormonal contraception related cancers, by sex.

ICD-O3 Categories	N (Female/ Male)	Infection related cancer (F/M)	Smoking related cancer (F/M)	Alcohol related cancer (F/M)	Reproductive / hormonal factors (F)
Anus (C21)	62/60	Case	Case	Control	Control
Bladder (C67)	31/79	Case	Case	Case	Control
Bone (C40-41)	60/100	Control	Control	Control	Control
Brain (C71)	29/29	Control	Control	Control	Control
Breast (C50)	5028/64	Control	Unclear[£]/Control	Case/Unclear[£]	Case
Cervix (C53)	5267/NA	Case/NA	Case/NA	Case/NA	Case
CNS (C72)	4/3	Control	Control	Control	Control
Colon (C18-20)	476/518	Control	Case	Case	Control
Endocrine gland (C75)	23/16	Control	Control	Control	Control
Endometrium (C54-55)	516/NA	Control/NA	Control/NA	Control/NA	Case
Eye and Adnexa (C69)	59/51	Case	Control	Control	Control
Fallopian tube (C57.0)	8/NA	Control/NA	Control/NA	Control/NA	Case
HL (C81)	155/168	Case	Control	Control	Control
Kaposi Sarcoma (C46)	690/881	Case	Control	Control	Case
Kidney (C64)	34/44	Control	Case	Case	Control
Larynx (C32)	42/353	Control	Case	Case	Control
Leukaemia (not Myeloid) (C91-95)	111/135	Case	Control	Control	Control
Liver (C22)	78/185	Case	Case	Case	Case
Lung Cancer (C33-34)	266/861	Control	Case	Case	Control
Melanoma (C43)	67/44	Control	Control	Control	Control
Meninges (C70)	13/5	Control	Control	Control	Control
Myeloid Leukaemia (ICD-10 C92)	160/171	Control	Case	Control	Control
Myeloma (C90)	180/180	Control	Control	Control	Control
Nasal cavity and nasopharynx (C11,C30,C31)	82/147	Case	Case	Unclear[£]	Control
NHL (C82-83)	425/494	Case	Control	Control	Control
Oesophagus (C15)	589/1008	Control	Case	Case	Control
Oral cavity and pharynx (C00-10, C12-14)	270/861	Case	Case	Case	Control
Ovaries (C56)	474/NA	Control/NA	Case/NA	Control/NA	Case
Pancreas (C25)	95/133	Control	Case	Case	Control
Penis (C60)	NA/49	NA/Case	NA/Case	NA/Control	NA
Peripheral nerves & ANS (C47)	2/7	Control	Control	Control	Control
Peritoneum and retroperitoneum (C48)	20/9	Control	Control	Control	Control
Placenta (C58.9)	77/NA	Control/NA	Control/NA	Control/NA	Unclear#
Prostate (C61)7	NA/719	NA/Control	NA/Control	NA/Control	NA
Scrotum (C63.2)	NA/7	NA/Control	NA/Control	NA/Control	NA
Small Intestine (C17)	12/11	Control	Control	Control	Control
Soft Tissue Sarcoma (C49)	140/148	Unclear[#]/Control	Control	Control	Control

Squamous cell carcinoma (C44)	78/74	Unclear[ε]/Case	Control	Control	Control
Stomach (C16)	168/256	Case	Case	Case	Control
Testes (C62)	NA/30	NA/Control	NA/Control	NA/Control	NA
Thymus (C37)	5/10	Control	Control	Control	Control
Thyroid (C73)	55/13	Control	Control	Control	Case
Vagina (C52)	59/NA	Case/NA	Case/NA	Case/NA	Case
Vulva (C51)	282/NA	Case/NA	Case/NA	Case/NA	Case

Decision to classify cancer types is based mainly on IARC Monographs on Carcinogenicity to Humans and other current literature available at the time of selection.[12] Cancer types with little research on their causality are defaulted to control status and tested using sensitivity analysis (see methods).
NA – Not applicable for sex specific cancers.
[ε]Lack of convincing evidence in current literature for cancer causality.

Table 3. Key results arising from case-control analyses of the JCS.

Exposure	Cancer type	Cases/controls[γ]	OR (95%CI)	Ref
HIV- / KSHV high titre	Kaposi Sarcoma	51/3293	12.0 (2.7-53.0)	[9]
HIV+ / KSHV high titre			1683 (545-5194)	
[$]HIV+	Kaposi Sarcoma	333/4399	47.1 (31.9-69.8)	[10]
	NHL	223	5.9 (4.3-8.1)	
	Cervix	1586	1.6 (1.3-2.0)	
	Hodgkin	154	1.6 (1-2.7)	
	Anogenital	157	2.2 (1.4-3.3)	
	Squamous cell skin	70	2.6 (1.4-4.9)	
	Oral cavity & pharynx	319	0.8 (0.5-1.3)	
	Liver	83	0.8 (0.4-1.7)	
	Lung	363	1.1 (0.7-1.6)	
Anti HBc+ & HBV DNA+ & HBsAg+	Hepatocellular	55/437	46.7 (21.0-103.9)	[11]
	M	/804		[25]
	F	/1370		
Current smoker 15g+ / day	Lung – M	33/58	23.9 (9.5-60.3)	
	Lung – F	6/11	50.9 (12.6-204.6)	
	Oesophageal – M	267	3.8 (2.3-6.1)	
	Oesophageal – F	138	3.1 (1.7-5.4)	
	Oral – M	87	7.5 (3.2-17.8)	
	Oral – F	37	3.9 (1.6-9.9)	
	Laryngeal – M	51	13.8 (3.0-63.9)	
Frequent alcohol drinker	Oesophageal - M	187	1.8 (1.2-2.8)	
	Oesophageal - F	38	1.7 (1.0-2.9)	
Smoke & alcohol	Oesophageal – M&F	265/546	4.4 (3.2-6.1)	
	M	/1383		[15]
	F	/2676		
Current smoker 15+g/day	Lung – M	115/95	37.4 (21.0-66.5)	
	Lung – F	9/20	18.5 (7.7-44.5)	
Non-electrical cooking	Lung – M	103/300	1.6 (1.2-2.2)	
	Lung- F	30/703	1.4 (0.8-2.2)	
Oral and/or injectable contraceptives		/1492		[16]
[ϕ]TSLU <10y	Breast	1664	1.66 (1.28-2.16)	
TSLU ≥10y			1.11 (0.91-1.36)	
TSLU <10y	Cervix	2182	1.38 (1.08-1.77)	
TSLU ≥10y			1.01 (0.84-1.22)	
TSLU <5y	Ovarian	135	0.69 (0.39-1.21)	
TSLU ≥5y			0.60(0.36-0.99)	
TSLU <5y	Endometrial	151	1.28 (0.71-2.32)	
TSLU ≥5y			0.44 (0.22-0.86)	

[γ]Number of controls listed once.

[$]Patients with unknown HIV status were excluded from the analysis.

[ϕ]TSLU=Time since last use.

M = Male; F= Female; y=years.